

Performability analysis of the second order semi-Markov chains in state and duration for wind speed modeling

Guglielmo D'Amico^a, Filippo Petroni^b, Flavio Prattico^c

^a*Dipartimento di Scienze del Farmaco, Università G. d'Annunzio, 66013 Chieti, Italy
E-mail: g.damico@unich.it*

^b*Dipartimento di Scienze Economiche e Aziendali, Università degli studi di Cagliari,
09123 Cagliari, Italy
E-mail: fpetroni@unica.it*

^c*Dipartimento di Ingegneria Meccanica, Energetica e Gestionale, Università degli studi
dell'Aquila, 67100 L'Aquila, Italy
E-mail: flavioprattico@gmail.com*

Abstract

In this paper second order semi-Markov reward models are presented and equations for the higher order moments of the reward process are presented for the first time and applied to wind energy production. A real application is executed by considering a database, freely available from the web, in which are included wind speed data taken from L.S.I. - Lastem station (Italy) and sampled every 10 minutes. We compute the expected total energy produced by using the blade Aircon HAWT - 10 kW.

Keywords: semi-Markov chains, synthetic time series, autocorrelation

1. Introduction

Nowadays the subject of wind speed modeling is becoming increasingly important. Wind power generation companies are intensely interested in quantifying the energy which can be extracted from wind by using different types of wind turbines in a given location.

Models to predict future wind speed and power production have been proposed extensively in last two decades. In general the wind speed is divided into a finite number of states. Transitions between states occur randomly in time. The installed blade produces energy depending on its technical characteristic and on the wind speed status.

Markov chains have been extensively used to model the behavior of wind

speed data. Several authors have discussed the use of techniques of Markov processes in wind speed modeling, see e.g., Shamshad et al. (2005), Nfaoui et al. (2004) and Youcef et al. (2003). The Markovian assumption has, especially in the modeling of wind speed, several flaws. In discrete time, waiting times in a state before making a transition to another state are geometrically distributed and consequently the memoryless property applies. This leads to a great simplification of the model which is unable to reproduce correctly the statistical properties of the real wind speed process.

Semi-Markov chains do not have this constraint, because the waiting time distribution function in the states can be of any type and this allow the data to speak for themselves without any restriction. D’Amico et al. (2011) was the first paper where semi-Markov chains were applied in the modeling of wind speed. In that paper first and second order semi-Markov models were proposed with the aim of generate reliable synthetic wind speed data. It was shown that all the semi-Markov models perform better than the Markov chain model in reproducing the statistical properties of wind speed data. In particular, the model recognized as being the more suitable is the second order semi-Markov model in state and duration.

One of the purposes of this paper is to provides methods for computing the accumulated energy produced by a blade in a temporal interval $[0, T]$. To this end we introduce semi-Markov reward processes.

Semi-Markov reward processes were applied in several domains, for example De Dominicis and Manca (1986) applied non-homogeneous semi-Markov reward processes to insurance disability problems. In Stenberg et al. (2007) backward semi-Markov reward processes were considered for calculating any integer moment of the reward process.

In this paper we give a generalization of these results by defining the second order semi-Markov reward process in state and duration and giving relations for computing the higher order moments of this process. We apply the theoretical results in computing the accumulated energy produced by a blade during a bounded time interval. The expected total energy produced gives important information on the feasibility of the investment in a wind farm and the riskiness of the investment can be measured in terms of variance, skewness, and kurtosis of the reward process. Finally notice that the technological characteristics of different blades are captured by the permanence reward and consequently we are able to choose among different blades to be installed at a given location. Additionally, we propose to employ a matrix notation that makes calculations easier and also provides a compact

form for equations of moments of the reward process.

2. The second order semi-Markov chain in state and duration

In this section we describe briefly the second order semi-Markov chain in state and duration, see D'Amico et al. (2012) for additional results.

Let consider a finite set of states $E = \{1, 2, \dots, S\}$ in which the system can be into and a complete probability space (Ω, F, \mathbb{P}) on which we define the following random variables:

$$J_n : \Omega \rightarrow E, \quad T_n : \Omega \rightarrow \mathbb{N}. \quad (1)$$

They denote the state occupied at the n -th transition and the time of the n -th transition, respectively. To be more concrete, by J_n we denote the wind speed at the n -th transition and by T_n the time of the n -th transition of the wind speed.

We assume that

$$\begin{aligned} \mathbb{P}[J_{n+1} = j, T_{n+1} - T_n = t | \sigma(J_s, T_s), J_n = k, J_{n-1} = i, T_n - T_{n-1} = x, 0 \leq s \leq n] \\ = \mathbb{P}[J_{n+1} = j, T_{n+1} - T_n = t | J_n = k, J_{n-1} = i, T_n - T_{n-1} = x] := {}_x q_{i,k,j}(t). \end{aligned} \quad (2)$$

Relation (2) asserts that, the knowledge of the values $J_n, J_{n-1}, T_n - T_{n-1}$ suffices to give the conditional distribution of the couple $J_{n+1}, T_{n+1} - T_n$ whatever the values of the past variables might be. Therefore to make probabilistic forecasting we need the knowledge of the last two visited state and the duration time of the transition between them. For this reason we called this model a second order semi-Markov chains in state and duration.

It should be remarked that in the paper by Limnios and Oprisan (2003) were defined n th order semi-Markov chains in continuous time. Anyway the dependence was only on past states and not on durations.

The conditional probabilities

$${}_x q_{i,k,j}(t) = \mathbb{P}[J_{n+1} = j, T_{n+1} - T_n = t | J_n = k, J_{n-1} = i, T_n - T_{n-1} = x]$$

are stored in a matrix of functions $\mathbf{q} = ({}_x q_{i,k,j}(t))$ named the second order kernel (in state and duration). The element ${}_x q_{i,k,j}(t)$ represents the probability that next wind speed will be in speed j at time t given that the current wind speed is k and the previous wind speed state was i and the duration in wind speed i before of reaching wind speed k was equal to x units of time.

From the knowledge of the kernel we can define the cumulated second order kernel probabilities:

$$\begin{aligned} {}_xQ_{i,k,j}(t) &:= \mathbb{P}[J_{n+1} = j, T_{n+1} - T_n \leq t | J_n = k, J_{n-1} = i, T_n - T_{n-1} = x] \\ &= \sum_{s=1}^t {}_xq_{i,k,j}(s). \end{aligned} \tag{3}$$

The process $\{J_n\}$ is a second order Markov chain with state space E and transition probability matrix ${}_x\mathbf{P} = {}_x\mathbf{Q}(\infty)$. We shall refer to it as the embedded Markov chain.

Define the unconditional waiting time distribution function in states k coming from state i with duration x as

$${}_xH_{i,k}(t) := \mathbb{P}[T_{n+1} - T_n \leq t | J_n = k, J_{n-1} = i, T_n - T_{n-1} = x] = \sum_{j \in E} {}_xQ_{i,k,j}(t). \tag{4}$$

The conditional cumulative distribution functions of the waiting time in each state, given the state subsequently occupied is defined as

$$\begin{aligned} {}_xG_{i,k,j}(t) &= \mathbb{P}[T_{n+1} - T_n \leq t | J_n = k, J_{n-1} = i, J_{n+1} = j, T_n - T_{n-1} = x] \\ &= \frac{1}{{}_xp_{i,k,j}} \sum_{s=1}^t {}_xq_{i,k,j}(s) \cdot 1_{\{{}_xp_{i,k,j} \neq 0\}} + 1_{\{{}_xp_{i,k,j} = 0\}} \end{aligned} \tag{5}$$

Define by $N(t) = \sup\{n : T_n \leq t\} \forall t \in \mathbb{N}$. We define the second order (in state and duration) semi-Markov chain as $Z(t) = (Z^1(t), Z^2(t)) = (J_{N(t)-1}, J_{N(t)})$.

For this model ordinary transition probability functions and transition probabilities with initial and final backward recurrence times were defined and computed D'Amico et al. (2012).

3. The Reward Model

In this section by following the line of research in Stenberg et al. (2007) we determine recursive equations for higher order moments of the second order semi-Markov reward chain in state and duration.

Let $\xi(t)$ denote the accumulated discounted reward during the time interval $(0, t]$ defined by the following relation,

$$\xi(t) = \sum_{0 < u \leq t} X_{N(u)-2} \psi_{J_{N(u)-2}, J_{N(u)-1}; J_{N(u)}}(B(u)) e^{-\delta u} \tag{6}$$

where: $B(u) = u - T_{N(u)}$ is the backward recurrence time process, $X_{N(u)} := T_{N(u)+1} - T_{N(u)}$ is the sojourn time in state $J_{N(u)}$ before the $N(u) + 1$ transition and $e^{-\delta}$ with $\delta \in [0, 1]$ is a one period deterministic discount factor.

The reward $\psi_{J_{N(u)-2}, J_{N(u)-1}; J_{N(u)}}(B(u))$ is more general than those considered in Stenberg et al. (2007) because at the same time it is state dependent on the current state $J_{N(u)}$ of the system; it depends on the last two visited states $J_{N(u)-2}, J_{N(u)-1}$; it is duration dependent in the current state because it is a function of the backward process $B(u)$ and finally it depends on the sojourn time in the past states being dependent on $X_{N(u)-2}$.

Let also denote by ${}_{x,v}\xi_{i,k}(t)$ the random variable, which has the distribution the same with the conditional distribution for the random variable $\xi(t)$ given that

$$J_{N(0)-1} = k, J_{N(0)-2} = i, B(T_{N(0)-1}) = v, X_{N(0)-2} = x$$

and let denote by ${}_{x,v}V_{i,k}^{(n)}(t) := E[({}_{x,v}\xi_{i,k}(t))^n]$.

In order to propose a matrix notation that simplifies calculations and provides a compact form for next equations we need to introduce the adopted notation and products.

Given two $m \times n$ matrices \mathbf{A} and \mathbf{B} , their Hadamard matrix product \boxtimes gives the $m \times n$ matrix C whose generic element is given by:

$$c_{ij} = a_{ij}b_{ij}.$$

Let \mathbf{A} be a $m^2 \times m$ matrix and \mathbf{B} be a m^2 column vector, their \otimes matrix product gives the m^2 column vector whose elements, for all $i, k \in \{1, 2, \dots, m\}$ are expressed by

$$C_{(i-1) \cdot |m| + k} = \sum_{j=1}^m A_{(i-1) \cdot |m| + k, j} B_{(k-1) \cdot |m| + j, 1}.$$

The first order moment of the reward process ${}_{x,v}\xi_{i,k}(t)$ is computed in the following Theorem.

Theorem 1. *The first order moment of the second order semi-Markov chain in state and duration satisfies the following matrix equation:*

$$\begin{aligned} {}_{x,v}\mathbf{V}^{(1)}(t) &= {}_{x,v}\mathbf{D}(t) \boxtimes {}_{x,v}\tilde{\Psi}(t) + \sum_{s=1}^t ({}_{x,v}\mathbf{B}(s) \cdot 1_{|E|}) \boxtimes {}_{x,v}\tilde{\Psi}(s) \\ &+ \sum_{s=1}^t e^{-\delta s} {}_{x,v}\mathbf{B}(s) \otimes {}_{x+s,0}\mathbf{V}^{(1)}(t-s) \end{aligned} \quad (7)$$

where $\forall i, k \in E$

$$\begin{aligned}
{}_{x,v}\mathbf{V}^{(1)}(t) &= \left({}_{x,v}V_{(i-1) \cdot |E|+k}^{(1)}(t) \right) = \left({}_{x,v}V_{i,k}^{(1)}(t) \right), \\
{}_x\Psi(v+u) &= ({}_x\psi_{(i-1) \cdot |E|+k}(v+u)e^{-\delta u}) = ({}_x\psi_{i,k;k}(v+u)e^{-\delta u}), \\
{}_x\Psi(v+u) \text{ is } |E|^2 \times 1 \text{ and } {}_{x,v}\tilde{\Psi}(t) &= \sum_{u=1}^t {}_x\Psi(v+u) \text{ is } |E|^2 \times 1, \\
{}_{x,v}\mathbf{D}(t) &= ({}_{x,v}D_{(i-1) \cdot |E|+k}(t)) = ({}_{x,v}D_{i,k}(t)) = \left(\frac{1 - {}_xH_{i,k}(t+v)}{1 - {}_xH_{i,k}(v)} \right) \\
{}_{x,v}\mathbf{B}(s) &= ({}_{x,v}B_{(i-1) \cdot |E|+k,j}(s)) = \left(\frac{{}_xq_{i,k;j}(s+v)}{1 - {}_xH_{i,k}(v)} \right)
\end{aligned}$$

and

$$\mathbf{1}_{|E|}^T = [1, 1, \dots, 1]^T.$$

Proof: Let consider the random variable ${}_{x,v}\xi_{i,k}(t)$. The time of next transition $T_{N(0)+1}$ can be greater of t or not. Consequently it results that:

$${}_{x,v}V_{i,k}^{(1)}(t) := E[{}_{x,v}\xi_{i,k}(t)] = E[{}_{x,v}\xi_{i,k}(t)1_{\{T_{N(0)+1} > t\}}] + E[{}_{x,v}\xi_{i,k}(t)1_{\{T_{N(0)+1} \leq t\}}]. \quad (8)$$

In the case $T_{N(0)+1} > t$ we have that

$${}_{x,v}\xi_{i,k}(t) = \sum_{u=1}^t {}_x\psi_{i,k;k}(u+v)e^{-\delta u} \quad (9)$$

and this event occurs with probability

$$\begin{aligned}
&\mathbb{P}(T_{N(0)+1} > t | T_{N(0)+1} > 0, T_{N(0)} = -v, J_{N(0)} = k, T_{N(0)-1} = -v-x, J_{N(0)-1} = i) \\
&= \frac{\mathbb{P}(T_{N(0)+1} > t, T_{N(0)+1} > 0, T_{N(0)} = -v | J_{N(0)} = k, T_{N(0)-1} = i, X_{N(0)-1} = x)}{\mathbb{P}(T_{N(0)+1} > 0, T_{N(0)} = -v | J_{N(0)} = k, T_{N(0)-1} = i, X_{N(0)-1} = x)} \\
&= \frac{\mathbb{P}(X_{N(0)} > t+v | J_{N(0)} = k, T_{N(0)-1} = i, X_{N(0)-1} = x)}{\mathbb{P}(X_{N(0)} > v | J_{N(0)} = k, T_{N(0)-1} = i, X_{N(0)-1} = x)} \\
&= \frac{1 - {}_xH_{i,k}(t+v)}{1 - {}_xH_{i,k}(v)} = {}_{x,v}D_{i,k}(t).
\end{aligned} \quad (10)$$

Then it results that

$$\mathbb{E}[x, v \xi_{i,k}(t) 1_{\{T_{N(0)+1} > t\}}] = {}_{x,v} D_{i,k}(t) \sum_{u=1}^t {}_x \psi_{i,k;k}(u+v) e^{-\delta u}. \quad (11)$$

The right hand side of (11) can be expressed in matrix form as follows:

$${}_{x,v} \mathbf{D}(t) \boxtimes {}_{x,v} \tilde{\Psi}(t). \quad (12)$$

In the second case, when $T_{N(0)+1} \leq t$, if we consider the next visited state $J_{N(0)+1}$ and the time of next transition $T_{N(0)+1}$ we have:

$${}_{x,v} \xi_{i,k}(t) = \left(\sum_{s'=1}^{T_{N(0)+1}} {}_x \psi_{i,k;k}(v+s') e^{-\delta s'} + {}_{v+T_{N(0)+1},0} \xi_{k,J_{N(0)+1}}(t-s) e^{-\delta T_{N(0)+1}} \right). \quad (13)$$

The event $\{J_{N(0)+1} = j, T_{N(0)+1} = s\}$ occurs with probability

$$\begin{aligned} & \mathbb{P}(J_{N(0)+1} = j, T_{N(0)+1} = s | T_{N(0)+1} > 0, T_{N(0)} = -v, J_{N(0)} = k, T_{N(0)-1} = -v-x, J_{N(0)-1} = i) \\ &= \frac{\mathbb{P}(J_{N(0)+1} = j, T_{N(0)+1} = s, T_{N(0)+1} > 0, T_{N(0)} = -v | J_{N(0)} = k, T_{N(0)-1} = i, X_{N(0)-1} = x)}{\mathbb{P}(T_{N(0)+1} > 0, T_{N(0)} = -v | J_{N(0)} = k, T_{N(0)-1} = i, X_{N(0)-1} = x)} \\ &= \frac{\mathbb{P}(J_{N(0)+1} = j, X_{N(0)} = s+v | J_{N(0)} = k, T_{N(0)-1} = i, X_{N(0)-1} = x)}{\mathbb{P}(X_{N(0)} > v | J_{N(0)} = k, T_{N(0)-1} = i, X_{N(0)-1} = x)} \\ &= \frac{{}_x q_{i,k;j}(s+v)}{1 - {}_x H_{i,k}(v)} = {}_{x,v} B_{(i-1) \cdot |E| + k, j}(s). \end{aligned} \quad (14)$$

Notice that the random variable ${}_{v+T_{N(0)+1},0} \xi_{k,J_{N(0)+1}}(t-s)$ is independent of the distribution of the joint random variable $(J_{N(0)+1}, T_{N(0)+1})$ because the accumulation process has the Markov property at transition times and consequently once the state $J_{N(0)+1}$ and the $T_{N(0)+1}$ are known its behaviour doesn't depends on the distribution of $(J_{N(0)+1}, T_{N(0)+1})$. Then by taking the expectation in (13) we get

$$\begin{aligned} & \mathbb{E}[{}_{x,v} \xi_{i,k}(t) 1_{\{T_{N(0)+1} \leq t\}}] \\ &= \sum_{j \in E} \sum_{s=1}^t \frac{{}_x q_{i,k;j}(s+v)}{1 - {}_x H_{i,k}(v)} \cdot \sum_{s'=1}^s {}_x \psi_{i,k;k}(v+s') e^{-\delta s'} \\ &+ \sum_{j \in E} \sum_{s=1}^t \frac{{}_x q_{i,k;j}(s+v)}{1 - {}_x H_{i,k}(v)} \cdot {}_{v+s,0} V_{k,j}^{(1)}(t-s) e^{-\delta s}. \end{aligned} \quad (15)$$

The right hand side of (15) can be expressed in matrix form as follows:

$$\begin{aligned} {}_{x,v}\mathbf{V}^{(1)}(t) &= \sum_{s=1}^t ({}_{x,v}\mathbf{B}(s) \cdot \mathbf{1}_{|E|}) \boxtimes {}_{x,v}\tilde{\mathbf{\Psi}}(s) \\ &\quad + \sum_{s=1}^t e^{-\delta s} {}_{x,v}\mathbf{B}(s) \otimes {}_{x+s,0}\mathbf{V}^{(1)}(t-s). \end{aligned} \quad (16)$$

A substitution of (12) and (16) in (8) concludes the proof.

By using similar techniques it is possible to get recursive equations for the higher order moments of the reward process.

First note that

$$\begin{aligned} {}_{x,v}V_{i,k}^{(n)}(t) &:= \mathbb{E}[({}_{x,v}\xi_{i,k}(t))^n] \\ &= E[({}_{x,v}\xi_{i,k}(t))^n \mathbf{1}_{\{T_{N(0)+1} > t\}}] + E[({}_{x,v}\xi_{i,k}(t))^n \mathbf{1}_{\{T_{N(0)+1} \leq t\}}]. \end{aligned} \quad (17)$$

In the case $T_{N(0)+1} > t$ we have that

$$({}_{x,v}\xi_{i,k}(t))^n = \left(\sum_{u=1}^t {}_x\psi_{i,k;k}(u+v) e^{-\delta u} \right)^n \quad (18)$$

and this event occurs with probability ${}_{x,v}D_{i,k}(t)$, see (10). Consequently it results that

$$E[{}_{x,v}\xi_{i,k}^{(n)}(t) \mathbf{1}_{\{T_{N(0)+1} > t\}}] = {}_{x,v}D_{i,k}(t) \left(\sum_{u=1}^t {}_x\psi_{i,k;k}(u+v) e^{-\delta u} \right)^n. \quad (19)$$

In the opposite case, when $T_{N(0)+1} \leq t$, we have that

$$\begin{aligned} {}_{x,v}\xi_{i,k}^{(n)}(t) &= \left(\sum_{s'=1}^{T_{N(0)+1}} {}_x\psi_{i,k;k}(v+s') e^{-\delta s'} + {}_{v+T_{N(0)+1},0}\xi_{k,J_{N(0)+1}}(t-s) e^{-\delta T_{N(0)+1}} \right)^n \\ &\quad \left(\sum_{s'=1}^{T_{N(0)+1}} {}_x\psi_{i,k;k}(v+s') e^{-\delta s'} \right)^n + {}_{v+T_{N(0)+1},0}\xi_{k,J_{N(0)+1}}^{(n)}(t-s) e^{-n\delta T_{N(0)+1}} \\ &\quad + \sum_{l=1}^{n-1} \binom{n}{l} \left(\sum_{s'=1}^s {}_x\psi_{i,k;k}(v+s') e^{-\delta s'} \right)^{n-l} \left({}_{v+T_{N(0)+1},0}\xi_{k,J_{N(0)+1}}^{(l)}(t-s) e^{-l\delta T_{N(0)+1}} \right). \end{aligned} \quad (20)$$

The event $\{J_{N(0)+1} = j, T_{N(0)+1} = s\}$ occurs with probability ${}_{x,v}B_{(i-1) \cdot |E| + k, j}(s)$.

Then, by using the already mentioned independence between ${}_{v+T_{N(0)+1}, 0}\xi_{k, J_{N(0)+1}}(t-s)$ and the joint random variable $(J_{N(0)+1}, T_{N(0)+1})$, by taking the expectation in (20) we get:

$$\begin{aligned}
& \mathbb{E}[{}_{x,v}\xi_{i,k}^{(n)}(t)1_{\{T_{N(0)+1} \leq t\}}] = \\
& + \sum_{j \in E} \sum_{s=1}^t \frac{{}_x q_{i,k;j}(s+v)}{1 - {}_x H_{i,k}(v)} \cdot \left(\sum_{s'=1}^s {}_x \psi_{i,k;k}(v+s')e^{-\delta s'} \right)^n \\
& + \sum_{j \in E} \sum_{s=1}^t \frac{{}_x q_{i,k;j}(s+v)}{1 - {}_x H_{i,k}(v)} \cdot {}_{v+s,0}V_{k,j}^{(n)}(t-s)e^{-\delta sn} \\
& + \sum_{j \in E} \sum_{s=1}^t \sum_{l=1}^{n-1} \frac{{}_x q_{i,k;j}(s+v)}{1 - {}_x H_{i,k}(v)} \binom{n}{l} \left(\sum_{s'=1}^s {}_x \psi_{i,k;k}(v+s')e^{-\delta s'} \right)^{n-l} \\
& \cdot {}_{v+s,0}V_{k,j}^{(l)}(t-s)e^{-\delta sl}.
\end{aligned} \tag{21}$$

If we substitute (19) and (21) in (17) and we represent the resulting expression in matrix form we obtain the following equation:

$$\begin{aligned}
{}_{x,v}V^{(n)}(t) &= {}_{x,v}D(t) \boxtimes {}_{x,v}\tilde{\Psi}^{(n)}(t) + \sum_{s=1}^t ({}_{x,v}B(s) \cdot 1_{|E|}) \boxtimes {}_{x,v}\tilde{\Psi}^{(n)}(s) \\
&+ \sum_{s=1}^t e^{-\delta sn} {}_{x,v}B(s) \otimes {}_{v+s,0}V^{(n)}(t-s) \\
&+ \sum_{s=1}^t \sum_{l=1}^{n-1} \binom{n}{l} {}_{x,v}\Psi^{(n-l)}(s) \boxtimes (e^{-\delta sl} {}_{x,v}B(s) \otimes {}_{v+s,0}V^{(l)}(t-s)).
\end{aligned} \tag{22}$$

It should be remarked that if $\forall i \in E$ and $\forall x \in \mathbb{N}$ we have that

$${}_x q_{i,k,j}(t) = q_{k,j}(t), \quad {}_x \psi_{i,k,j}(t) = \psi_{k,j}(t)$$

then the second order semi-Markov reward chain model in state and duration collapses in a standard semi-Markov reward chain model and we recover exactly the results by Stenberg et al. (2007).

4. Application to real data

To check the validity of our model we perform a comparison of the behavior of real data and wind speeds generated through Monte Carlo simulations

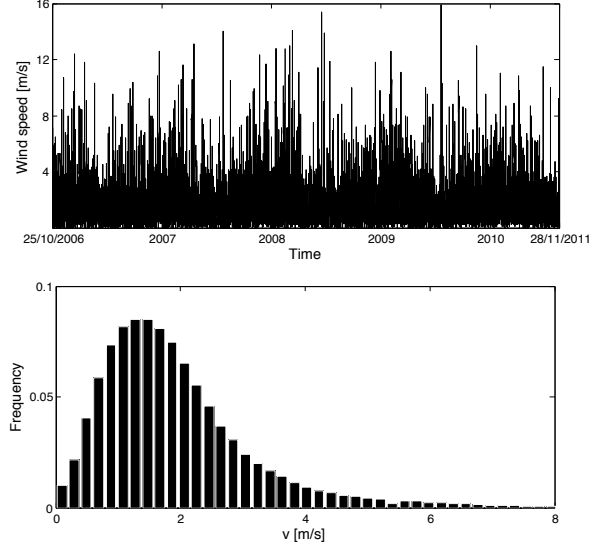


Figure 1: Time series of wind speed and its empirical distribution.

based on the models described above. In this section we describe the database of real data used for the analysis, the method used to simulate synthetic wind speed time series and, at the end, we compare results from real and simulated data.

The data used in this analysis are freely available from [http : //www.lsi-lastem.it/meteo/page/dwnldata.aspx](http://www.lsi-lastem.it/meteo/page/dwnldata.aspx). The station of L.S.I. -Lastem is situated in Italy at N 45 28' 14,9" – E 9 22' 19,9" and at 107 *m* of altitude. The station use a combined speed-direction anemometer at 22 *m* above the ground. It has a measurement range that goes from 0 to 60 *m/s*, a threshold of 0,38 *m/s* and a resolution of 0,05 *m/s*. The station processes the speed every 10 minute in a time interval ranging from 25/10/2006 to 28/06/2011. During the 10 minutes are performed 31 sampling which are then averaged in the time interval. In this work, we use the sampled data that represents the average of the modulus of the wind speed (*m/s*) without a considered specific direction. The database is then composed of about 230thousands wind speed measures ranging from 0 to 16 *m/s*. The time series, together with its empirical probability density distribution are represented in Figure 1. The lower panel of the same figure shows that the wind speed follows a Weibul distribution.

To be able to model the wind speed as a semi-Markov chain the state space of wind speed has been discretized. In the example shown in this work we discretized wind speed into 8 states chosen to cover all the wind speed distribution. The state space is numerically represented by the set $E = \{0 - 1, 1 - 2, 2 - 3, 3 - 4, 4 - 5, 5 - 6, 6 - 7, > 7\}(m/s)$. From the discretized wind speeds we estimated the probabilities \mathbf{P} and G to generate synthetic trajectories for a second order semi-Markov model in state and duration.

We give here the Monte Carlo algorithm used to simulate the trajectory in the time interval $[0, T]$ where T is equal to the time length of the real data. The output of the algorithm consists in the successive visited states $\{J_0, J_1, \dots\}$ and the jump times $\{T_0, T_1, \dots\}$ up to the time T . Denote by $X_n = T_{n+1} - T_n$.

- 1 Set $n = 0$, $J_{-1} = i$, $J_0 = k$, $X_{-1} = x$, horizon time $= T$;
- 2 Sample J from $_{X_{n-1}}p_{J_{n-1}J_n}(\cdot)$ and set $J_{n+1} = J(\omega)$;
- 3 Sample W from $_{X_{n-1}}G_{J_{n-1}, J_n, J_{n+1}}(\cdot)$ and set $T_{n+1} = T_n + W(\omega)$;
- 4 If $T_{n+1} \geq T$ stop, else set $X_n = T_{n+1} - T_n$ and $n = n + 1$ and go to 2.

We show in Figure 2 a short sample of 160 hours of the real and simulated trajectories just for comparison reason. We apply our model to a real case of energy production. For this reason we choose a commercial wind turbine, a 10 kW Aricon HAWT with a power curve given in Figure 3. The power curve of a wind turbine represents how it produces energy as a function of the wind speed. In this case we have no production of energy in the interval 0-2 m/s , the wind turbine produces energy linearly from 3 m/s to 10 m/s , than, with increasing wind speed the production remain constant until the limit of wind speed in which the wind turbine is stopped for structural reason. Through this power curve we transform the wind speed of our real and synthetic data into the equivalent energy produced at each different state of wind speed, in order to validate our model in a real case of energy production. For this reason, to check the proper functioning of the proposed model, we do several comparisons between real and synthetic data. In Figure 4 is showed the daily energy production of real and synthetic data. Another comparison between real and synthetic data is made by the cumulated energy produced in a time interval. In Figure 5 are showed these values as a function of time. The

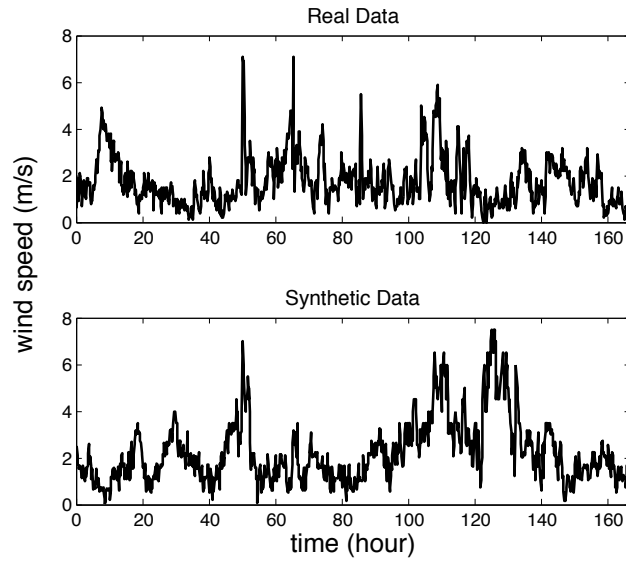


Figure 2: Comparison of the time series of wind speed from real and simulated data

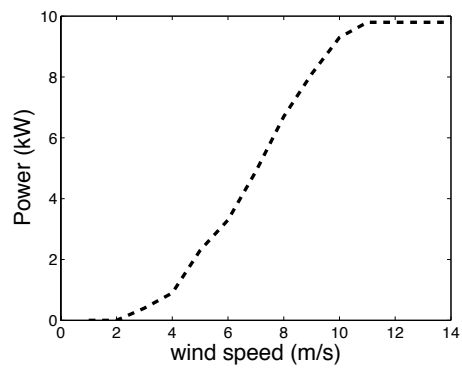


Figure 3: Power curve of the 10 kW Aricon HAWT wind turbine.

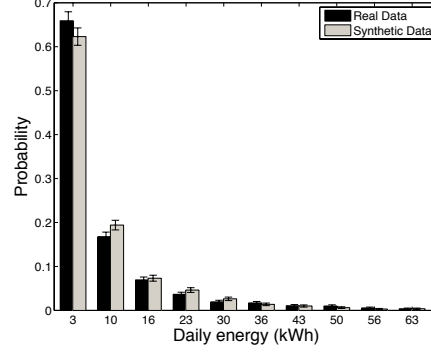


Figure 4: Comparison of the daily energy production for the real and simulated data. The error bar represent one standard deviation.

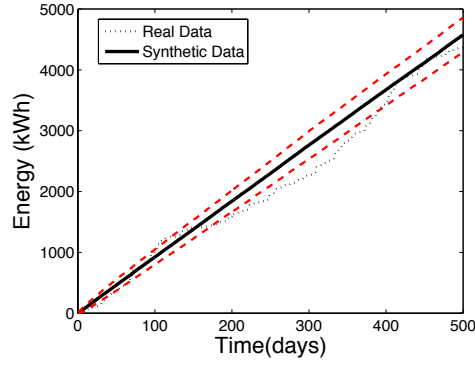


Figure 5: Comparison between the cumulated energy produced by real and simulated data.

two dashed line represents the standard deviation of the synthetic data. It is interesting to note that the trend of the real cumulated energy produced remains almost inside the area formed by the two dashed line. We generate 500 different trajectories and for each one we calculate the cumulated energy for all the length of the time series. In Figure 6 is plotted the distribution of the cumulated energy produced by the different trajectories. The black point represents the energy produced by the real data. The figure shows that simulated and real data give the same value for produced energy in the time interval within statistical error.

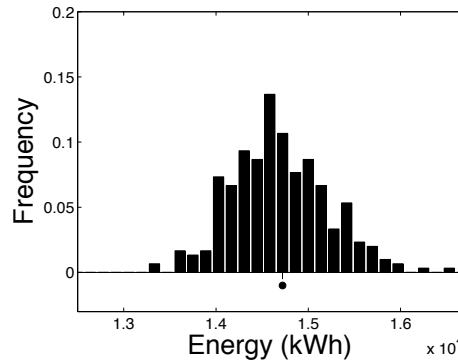


Figure 6: Distribution of the cumulated energy produced by the real and simulated trajectories.

5. Conclusion

The wind is a very unstable phenomenon characterized by a sequence of lulls and sustained speeds, and a good wind generator must be able to reproduce such sequences. We have then modeled wind speed through a second order semi-Markov model. Our work is motivated by the presence of persistence in the wind speed process. The purposes of this paper is to provides theoretical methods for computing the accumulated energy produced by a blade in a temporal interval $[0, T]$. We have shown, by means of Monte Carlo simulations, that the proposed model is able to reproduce well the behavior of real data as far as energy production from wind speed is concerned.

References

References

- [1] V. Barbu, M. Boussemart and N. Limnios, Discrete Time Semi-Markov Model for Reliability and Survival Analysis, *Commun Stat Theory Methods*, **33** (2004) 2833-2868.
- [2] Blasi, A., Janssen, J., Manca, R. (2004). Numerical treatment of homogeneous and non-homogeneous semi-Markov reliability models. *Communications in Statistics, Theory and Methods*, v. 33, 697-714.
- [3] E. Çinlar, Markov renewal theory, *Adv Appl Probab*, **1** (1969) 123-187.

- [4] G. D'Amico, J. Janssen, and R. Manca, European and American options: The semi-Markov case, *Physica A*, **388** (2009) 3181-3194.
- [5] G. D'Amico, Guillen, M., and R. Manca, Full backward nonhomogeneous semi-Markov processes for disability insurance models: a Catalunya real data application, *Insurance: Mathematics and Economics*, **45** (2009) 173-179.
- [6] G. D'Amico, J. Janssen, and R. Manca, Initial and Final Backward and Forward discrete time non-homogeneous semi-Markov Credit Risk Models, *Methodology and Computing in Applied Probability*, **12**(2) (2010), 215 - 225.
- [7] G. D'Amico, F. Petroni, and F. Prattico, First and second order semi-Markov chains for wind speed modeling, *Submitted* (2012), [arXiv:1206.2452](https://arxiv.org/abs/1206.2452).
- [8] J. Janssen and R. Manca, *Semi-Markov Risk Models for Finance, Insurance and Reliability* Springer (2007).
- [9] N. Limnios and G. Opreşan, *Semi-Markov Processes and Reliability*. Birkhäuser, Boston (2001).
- [10] N. Limnios, G. Opreşan. *An introduction to Semi-Markov Processes with Application to Reliability*, In D.N. Shanbhag and C.R. Rao, eds., *Handbook of Statistics*, **21**, (2003), 515-556.
- [11] H. Nfaoui, H. Essiarab, A.A.M. Sayigh, A stochastic Markov chain model for simulating wind speed time series at Tangiers, Morocco, *Renewable Energy* **29** (2004) 1407-1418.
- [12] A. Shamshad, M.A. Bawadi, W.M.W. Wan Hussin, T.A. Majid, S.A.M. Sanusi, First and second order Markov chain models for synthetic generation of wind speed time series, *Energy* **30** (2005) 693-708.
- [13] F. Youcef Ettoumi, H. Sauvageot, A.-E.-H. Adane, Statistical bivariate modeling of wind using first-order Markov chain and Weibull distribution, *Renewable Energy* **28** (2003) 1787-1802.
- [14] F. Stenberg, R. Manca, D. Silvestrov, An algorithmic approach to discrete time non-homogeneous backward semi-Markov reward processes

with an application to disability insurance. *Methodology and Computing in Applied Probability* **9** (2007) 497-519.